

Common polymorphic variation in the genetically diverse African insulin gene and its association with size at birth

Clive J. Petry · Pura Rayco-Solon · Anthony J. C. Fulford ·
John D. H. Stead · Dianne L. Wingate · Ken K. Ong ·
Giorgio Sirugo · Andrew M. Prentice · David B. Dunger

Received: 29 January 2009 / Accepted: 30 April 2009 / Published online: 12 May 2009
© The Author(s) 2009. This article is published with open access at Springerlink.com

Abstract The insulin variable number of tandem repeats (*INS* VNTR) has been variably associated with size at birth in non-African populations. Small size at birth is a major determinant of neonatal mortality, so the *INS* VNTR may influence survival. We tested the hypothesis, therefore, that genetic variation around the *INS* VNTR in a rural Gambian population, who experience seasonal variation in nutrition and subsequently birth weight, may be associated with foetal and early growth. Six polymorphisms flanking the *INS* VNTR were genotyped in over 2,500 people. Significant associations were detected between the maternally inherited SNP 27 (rs689) allele and birth length [effect size 17.5

(5.2–29.8) mm; $P = 0.004$; $n = 361$]. Significant associations were also found between the maternally inherited African-specific SNP 28 (rs5506) allele and post-natal weight gain [effect size 0.19 (0.05–0.32) z score points/year; $P = 0.005$; $n = 728$]. These results suggest that in the Gambian population studied there are associations between polymorphic variation in the genetically diverse *INS* gene and foetal and early growth characteristics, which contribute to overall polygenic associations with these traits.

Abbreviations

<i>IGF2</i>	Insulin-like growth factor 2 gene
<i>INS</i>	Insulin gene
LBW	Low birth weight
qTDT	Quantitative transmission disequilibrium test
SGA	Small for gestational age
SNP	Single nucleotide polymorphism
VNTR	Variable number of tandem repeats

C. J. Petry, P. Rayco-Solon, and A. J. C. Fulford contributed equally to this work and should be considered as equal first authors.

Electronic supplementary material The online version of this article (doi:10.1007/s00439-009-0681-2) contains supplementary material, which is available to authorized users.

C. J. Petry (✉) · D. L. Wingate · K. K. Ong · D. B. Dunger
Department of Paediatrics, Box 116, Addenbrooke's Hospital,
University of Cambridge, Hills Road, Cambridge CB2 0QQ, UK
e-mail: cjp1002@cam.ac.uk

P. Rayco-Solon · A. J. C. Fulford · A. M. Prentice
MRC International Nutrition Group,
London School of Hygiene and Tropical Medicine,
London WC1E 7HT, UK

P. Rayco-Solon · A. J. C. Fulford · G. Sirugo · A. M. Prentice
MRC Keneba, Medical Research Council Laboratories,
P.O. Box 273, Banjul, The Gambia

Present Address:

G. Sirugo
Unità di Genetica Medica, Ospedale San Pietro and Dipartimento
di Biopatologia e Diagnostica per Immagini,
Università di Tor Vergata, Rome, Italy

J. D. H. Stead
Department of Genetics,
University of Leicester, Leicester LE1 7RH, UK

Present Address:

J. D. H. Stead
Institute of Neuroscience,
Carleton University, Ottawa K1S 5B6, Canada

K. K. Ong
MRC Epidemiology Unit,
Cambridge CB2 0QQ, UK

Introduction

According to the “Out of Africa” or “African Replacement” hypothesis every non-African human being is descended from a founder group of homo sapiens living in Africa, a small subset of whom at some point in history dispersed into the wider world and displaced earlier human species resident there (reported by Stringer and McKie 1996). This explains the greater genetic diversity found in sub-Saharan-African populations in comparison to other ethnicities (Taillon-Miller et al. 2000).

One example of a heritable region found to be more genetically diverse in sub-Saharan-Africans is the insulin (*INS*) variable number of tandem repeats (VNTR), a 11p15.5 chromosomal locus at which particular 14–15 base pair repeats are present in variable numbers in different individuals within a population (Stead and Jeffreys 2002; Stead et al. 2003). In African populations the *INS* VNTR has been divided into 22 lineages, whereas in non-African populations there are only three lineages (grouped as class I, and classes IIIA and IIIB, according the number of repeats in the VNTR). Deviation from Mendelian transmission ratios has been observed with these two major classes (classes I and III; Eaves et al. 1999), implicating this genetic region as conferring a survival advantage. This may operate through associations with size at birth (i.e. birth weight, birth length or head circumference at birth) since: (1) historically low birth weight (LBW) has been the major determinant of neonatal mortality (McCormick 1985), (2) genetic diversity that alters size at birth may therefore have influenced survival in the past (Dunger et al. 2007), and (3) in non-African populations the *INS* VNTR has been variably associated with size at birth and post-natal weight gain (Dunger et al. 1998; Ong et al. 2004; Mitchell et al. 2004; Bennett et al. 2004; Lindsay et al. 2003; Heude et al. 2006; Vu-Hong et al. 2006; Landmann et al. 2006; Osada et al. 2007; Mook-Kanamori et al. 2007). Inconsistent reports in different populations may reflect genetic heterogeneity and selection of cases. Opposite associations have been observed in different populations (Lindsay et al. 2003), which could reflect linkage with adjacent polymorphisms in imprinted cluster genes on chromosome 11 that have been variably associated with weight gain (Gu et al. 2002; Petry et al. 2005; Paquette et al. 1998). Examples of these include insulin-like growth factor 2 (*IGF2*), whose protein product IGF-II is very important in foetal growth, and is both exclusively paternally expressed in foetal life and located adjacent to the *INS* gene, and *H19*, an *IGF2* repressor which is exclusively maternally expressed during foetal life (Lustig et al. 1994).

Such genetic factors may interact with effects associated with other established factors related to size at birth and early weight gain, including pre-pregnancy maternal weight

and maternal weight gain during pregnancy, gestational age, gender and parity (Ong et al. 2002). Whilst in developed countries maternal macronutrient intake during pregnancy may (Moore et al. 2004) or may not (Mathews et al. 1999) contribute to size at birth, in less developed populations such relationships are better established. An example of this has been observed in The Gambia, where severe seasonal variation in nutrition leads to a drop in birth weight of around 250 g and an increased prevalence of babies born with a LBW (Ceesay et al. 1997). In such populations with varying amounts and qualities of nutrition, there are likely to have been intense genetic pressures historically to boost foetal and early growth and therefore chances of surviving the critical neonatal period. In light of this we chose to study this Mandinkan African population from The Gambia who still experience extreme environmentally related nutritional pressure which affects size at birth, post-natal weight gain and long term survival (Ceesay et al. 1997). As well as having the nutritional pressures on their foetal and early growth, they also have the increased genetic diversity (in comparison to non-African populations) of the African *INS* VNTR. We sought to test the hypothesis that genetic variation around the *INS* VNTR in this population is associated with size at birth and early growth.

Materials and methods

Population

This was a retrospective cohort study of all live births in three subsistence-farming villages of the West Kiang District in The Gambia from 1976 to the present. There is a seasonal agricultural system that revolves around an annual rainy season, which occurs from July to November. Regular ante- and post-natal care has been provided by resident midwives to all women of childbearing age since the mid-1970s. Birth weights have been recorded from three villages (Keneba, Manduar and Kantong Kunda) since 1976. They were recorded to the nearest 10 g by the trained midwives with a Salter spring balance and tared sling (Salter Industrial Measurements Ltd., West Bromwich, UK), which was calibrated regularly. LBW was defined as weight <2,500 g. Measurement of gestational age was introduced in 1978, which was assessed by using the physical and neurological scoring system validated by Dubowitz et al. (1970) by medical doctors within 5 days of delivery. Prematurity was defined as a gestational age <37 completed weeks. Small for gestational age (SGA) was defined as a birth weight less than the tenth percentile of gestational age based on the reference standard by Williams et al. (1982). Thick and thin blood smears for malaria routinely taken from women during 43,385 antenatal and well-women clinic visits were pos-

itive for malaria on 3,276 occasions between 1978 and 2003. For a more detailed description of the population and its surrounds see Rayco-Solon et al. (2004).

There were two previous nutritional supplementation trials, which included ≥ 1 of the three villages (Ceesay et al. 1997; Prentice et al. 1987). The first-order effects of supplementation (fitted with the use of a single indicator variable) on preterm births and on SGA were not significant. The seasonality of either the preterm births or SGA was not significantly modified by supplementation when season was modelled by a simple binary variable indicating the hungry season (July to November). Thus, we ignored supplementation in the rest of the analysis.

The recording of births and deaths in these villages predated the formation of the Joint Gambian Government/Medical Research Council Gambia Ethics Committee. Approval for the continuation of the demographic surveillance was granted when the committee was formed in 1981. Approval for the genetic component of the study was granted in 2003. Informed consent was gained from either each study participant or their parents, as appropriate.

DNA samples

DNA from this population was extracted from blood, buffy coats or peripheral blood mononuclear cells using standard methods as part of the establishment of the Gambian DNA biobank (Sirugo et al. 2004). *INS* single nucleotide polymorphism (SNP) genotyping was successfully performed on 2,613–2,652 individuals, with between 124 and 553 informative DNA trios (i.e. mother, father and offspring) depending on the genotype distribution of the SNP in question and the availability of recorded birth phenotype (i.e. length, weight and head circumference). For each of the SNPs between 97.9 and 99.4% of the samples were successfully genotyped. Repeat genotyping gave 99.0–99.7% concordance rates.

Genotyping strategy

To genotype all the SNPs close to the *INS* VNTR would have involved genotyping 57 SNPs (Stead et al. 2003) and

given our population size, would have meant that some *INS* haplotypes were present in less than 10 people. Hence we developed a strategy whereby genotyping a reduced number of SNPs would give us closely related haplotype groups (after reconstructing phases) containing individuals with closely related haplotypes (Stead and Jeffreys 2002). Using the published African haplotype prevalences (Stead et al. 2003) and by assuming that a DNA sample's assigned haplotype was correct if it contained the *INS* VNTR haplotype that was most prevalent within that group, it was found that 85.3% of samples could have their haplotypes correctly assigned by genotyping just 5 SNPs (SNPs 24, 27, 28, 34, 38) and 1 indel ('SNP' 69) [from Stead et al. (2003); Fig. 1]. This strategy was therefore chosen to optimise the balance between the amount of genotyping 'effort' (cost, amount of DNA to be used, etc.) that needed to be made and useful genetic information gained. A subset of 100 randomly chosen DNA samples were then genotyped for these SNPs to validate them in our population. SNP allele population prevalences did not differ from those previously published in different African populations (Supplementary Table SD1 and Stead and Jeffreys 2002) so we then proceeded to use this genotyping strategy across the whole population.

Genotyping of *INS* VNTR polymorphisms

SNPs 24 (rs3842738), 27 (rs689), 28 (rs5506), 38 (rs3842752) and 69 (rs3842740) were genotyped using the 5'-nuclease technique and standard protocols (Taqman, Assays by Design; Applied Biosystems, Warrington, Cheshire, UK). Primer and probe sequences are shown in Table 1. PCR reactions contained 4 ng genomic DNA, 1 \times TaqMan universal PCR master mix, forward and reverse primers (both 72 nM), and VIC- and FAM-labelled probes (both 16 nM) in a total reaction volume of 5 μ l. The reaction mixes were incubated at 50°C for 2 min and 95°C for 10 min followed by 40 cycles 95°C for 15 s and 60°C for 1 min. Completed PCRs were read on an ABI Prism 3100 Genetic Analyzer (Applied Biosystems) and analysed using their allelic discrimination sequence detection software.

Table 1 Primer and probe sequences used in the 5'-nuclease assays and PCR-RFLP assay (for SNP 34) used to genotype the *INS* SNPs

SNP	Forward primer sequence	Reverse primer sequence	VIC-labelled probe sequence	FAM-labelled probe sequence
24	caccaccggccctaatgg	ccgctggctttatagctcaga	cccctacctgtcaacc	cccctacctetcaacc
27	gggcacctggccttcag	ccatggcagaaggacagtga	cctgcctgtctcccaga	ctgcctgtcaccaga
28	agaggacctgcagggtga	ggagcgcaggagcag	atggcggttggc	atggcgagtggc
34	ggggagaagtactgggatca	gggacacaggagcacagt	N/A	N/A
38	ccagctggagaactactgcaa	gttcaaggctttatccatctct	ctgcgggctgcgt	cctgcaggctgcgt
69	ggacaggctgcatcagaagag	gcccacctgacgcaaag	caggctcttgcgtcca	aagcaggctgttcca

N/A not applicable

SNP 34 (rs3842748) was genotyped using PCR amplification and restriction fragment length polymorphism analyses. Oligonucleotide primer sequences are shown in Table 1. A measure of 10 ng of DNA was amplified along with $1 \times$ Bionline NH_4 buffer, 200 μM each dNTP, 1 mM magnesium, 6 pmol each oligonucleotide primer and 0.5 units Biotaq DNA polymerase (Bionline, London, UK) in a total reaction volume of 10 μl . The mix was incubated for 5 min at 94°C , followed by 19 cycles of 94°C for 45 s, 59°C for 45 s (dropping 0.5°C per cycle) and 72°C for 45 s. After this the mix underwent 15 cycles of 94°C for 45 s, 49°C for 45 s and 72°C for 45 s, followed by a final incubation at 72°C for 10 min. The PCR mix was then incubated with 1 unit *Dra*III (New England Biolabs, Hitchin, Herts., UK) at 37°C for 16 h. This produced a 223 bp band for the C allele and 118 and 105 bp bands corresponding to the G allele when the sample was separated by agarose gel electrophoresis.

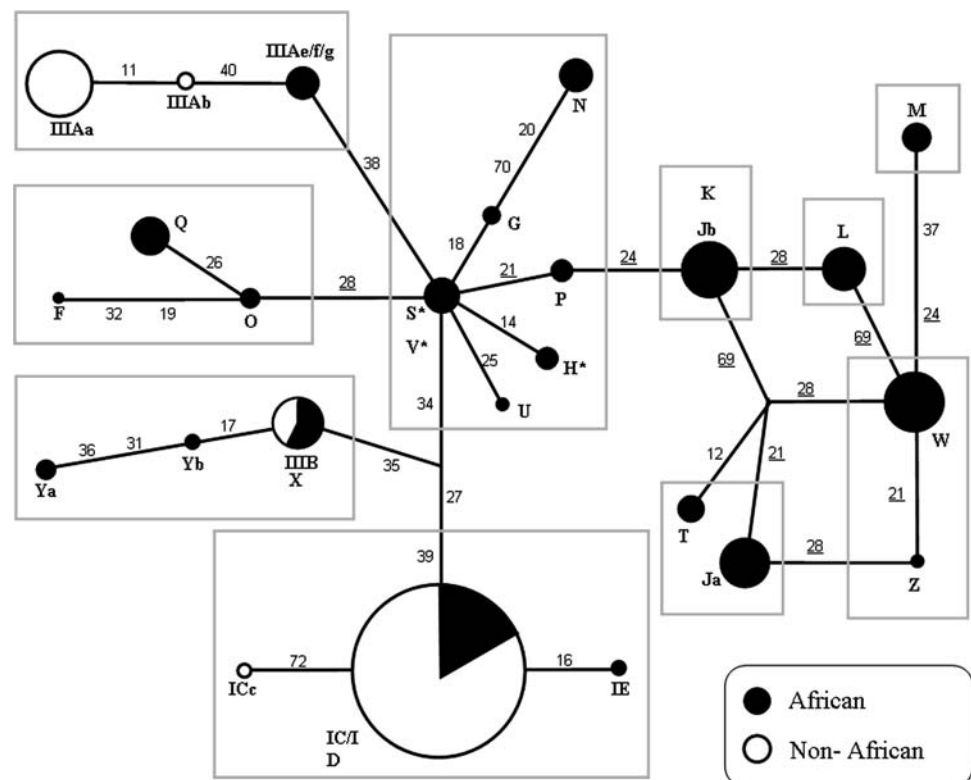
Statistical analyses

We examined separately the association of each SNP with four continuous outcomes: weight, length and head circumference at birth and weight gain in the first 2 years of life. Whilst straightforward least squares regression was used for the birth outcomes, weight gain analysis was performed in three stages. First weight-for-age z scores were calculated for each observation using age- and sex-specific means and standard deviations calculated from the Keneba

child clinic data. Next the slope and its standard error for the regression line of z score on age was calculated for each child. Finally these slopes were used as the outcome variable in a least squares regression on genotype. Only children who had at least 5 observations before 2 years of age and at least one in both the first and second years of life, and those for whom standard error for the regression slope of z score on age was less than 0.4, were included in this analysis.

Each SNP was fitted in regression analysis as a linear term: the effect of each copy of the minor allele was assumed to be additive. Both the maternal and infant genotypes were considered. We also considered parent-of-origin effects for the infant's genotype. Where these could be determined, we additionally fitted models with separate terms for alleles derived from mother and father. In order to test for a parent-of-origin effect these last models were reparameterised in terms of an interaction term for the difference between the effects of an allele according to whether it was inherited from the father or the mother. In order to examine the original class I/class III axis found to be predictive of birth size in non-African populations we constructed a variable computed from the difference between the number of copies of classes I and III haplotypes (defined by SNP 27, and SNPs 34 and 38, respectively; see Fig. 1). This variable therefore took values between -2 and 2 ; individuals without either of these haplotypes were uninformative and omitted from this analysis.

Fig. 1 Median joining network showing genetically related *INS* haplotypic groups in Mandinkan people from The Gambia, established by genotyping SNPs 24 (rs3842738), 27 (rs689), 28 (rs5506), 34 (rs3842748) and 38 (rs3842752), and indel 69 (rs3842740). SNP numbers and diagram adapted from Stead et al. (2003). Reproduced with permission from Cold Spring Harbour Laboratory Press © 2009, Stead et al. (2003)



The main analyses of SNPs presented here fitted a general linear model of association, with random effects to allow for intra-family correlation using the generalised least squares method (Stat Corp LP 2007). Secondary analyses for comparison purposes were fitted using a quantitative TDT (qTDT) model (Allison 1997), regressing outcome on genotype whilst controlling for mating type (AA × Aa, Aa × Aa or aa × Aa—other mating types are uninformative and hence excluded from this analysis). An extension of this method (George et al. 1999) was used to incorporate parent-of-origin effects. In order to allow for the inclusion of several children from each family, we again fitted a “between-family” random effect, as suggested by Allison (1997). Polygamy is very common in this society (children in this study had on average 2.4 full siblings and 2.6 half-siblings; 2.5 of which have a different mother) complicating the definition of nuclear families. Since children sharing the same mother are more closely related than those sharing the same father (due to exposure to both genes transmitted from the mother and the maternal intra-uterine environment), for the purposes of this analysis, we defined family members as children sharing the same mother. (We are aware that this is clearly imperfect since there will also be some dependence between other relatives. However, we are reassured that these could be ignored since even allowing for siblings did not profoundly affect conclusions.) Each analysis was repeated with and without the following covariates: sex of the child, village where the birth took place, year of birth (linear trend), season of birth (binary hungry versus harvest—first half of year vs. second) and parity (binary according to whether the infant was the mother’s first delivery). qTDT analysis was not appropriate for the class I/III variable so only the association models with random effects were fitted. All statistical

analysis was performed using Stata version 10 (Stat Corp LP 2007). For the association analyses [based on genotyping 2,700 samples, using the methods of Lange et al. (2002, 2004) and an additive model] for a fraction of the variance of birth weight of 1% the study had statistical power of 90.2–99.9 % ($\alpha = 0.05$ –0.001). Due to potential effects of multiple testing a Bonferroni correction was applied such that $P < 0.008$ was considered statistically significant throughout.

Results

General population characteristics

Table 2 shows the birth characteristics from the three villages from which our study population was drawn. Table 3 shows the genotype frequencies for the *INS* genotypes from our population, all of which were consistent with Hardy Weinberg equilibrium. Figure SD1 in the supplementary data section shows the linkage disequilibrium between the 5 *INS* SNPs and 1 indel that were genotyped in this population.

Simple associations of genotypes and allele transmission with birth and early growth characteristics

Table 4 shows the effect sizes of the associations between the various *INS* VNTR SNPs and the birth and early growth phenotypes. The supplementary data Table SD2 shows the effect sizes of the linkage and association of the transmission of these SNP alleles with phenotype using qTDT, when trying to confirm these associations. None of the associations were significantly modified by adjusting for the season of birth.

Table 2 Summary of the birth characteristics from each of the three villages in the West Kiang District in The Gambia from which the study population was drawn

Village	Kantong Kunda	Keneba	Manduar	Overall
	Proportion (<i>n</i>)	Proportion (<i>n</i>)	Proportion (<i>n</i>)	Proportion (<i>n</i>)
Sex ratio (boys/girls)	1:1.20 (266)	1:1.08 (913)	1:0.92 (322)	1:1.06 (1,501)
Primiparous	14.9% (261)	16.0% (898)	15.8% (310)	15.8% (1,469)
LBW	13.9% (266)	11.4% (913)	11.2% (322)	11.8% (1,501)
Premature	13.8% (218)	8.1% (755)	7.3% (247)	8.9% (1,220)
Born January–June	47.4% (266)	49.2% (910)	45.8% (321)	48.2% (1,497)
Village	Kantong Kunda	Keneba	Manduar	Overall
	Mean (SD; <i>n</i>)	Mean (SD; <i>n</i>)	Mean (SD; <i>n</i>)	Mean (SD; <i>n</i>)
Weight (kg)	2.90 (0.42; 266)	2.93 (0.40; 913)	2.98 (0.43; 322)	2.94 (0.41; 1,501)
Length (mm)	495.6 (26.2; 140)	489.9 (26.6; 405)	494.7 (27.9; 139)	492.0 (26.9; 684)
Head circumference (mm)	339.8 (15.8; 223)	338.6 (14.4; 727)	339.6 (15.7; 263)	339.0 (15.0; 1,213)
Gestational age (weeks)	38.8 (1.7; 229)	38.9 (1.6; 760)	39.2 (1.6; 257)	38.9 (1.6; 1,246)

Table 3 Numbers (and %) of individuals split by *INS* genotypes in the West Kiang District Gambian population. All the genotypes were consistent with Hardy Weinberg equilibrium

SNP	Polymorphism	No. of copies of non-ancestral allele			χ^2 -test for consistency with Hardy Weinberg equilibrium	
		0	1	2	χ^2	<i>P</i>
24	C → G	628 (23.8%)	1,305 (49.5%)	704 (26.7%)	0.23	0.89
27	T → A	1960 (74.0%)	636 (24.0%)	53 (2.0%)	0.03	0.99
28	C → T	1,528 (58.3%)	959 (36.6%)	134 (5.1%)	1.09	0.58
34	G → C	1,811 (69.3%)	745 (28.5%)	57 (2.2%)	3.75	0.15
38	C → T	2,082 (78.5%)	536 (20.2%)	34 (1.3%)	0.01	1.00
Indel 69	– → +ttgc	1,261 (47.8%)	1103 (41.8%)	275 (10.4%)	2.14	0.34

SNP 24 was not associated with any of the foetal and early growth phenotype. However, SNP 27 showed a significant association between birth length and infant's allele transmitted from the mother controlled for maternal genotype ($P = 0.004$; by qTDT $P = 0.015$).

Significant associations shown by SNP 28 were only evident with post-natal growth, where maternal transmission of the T allele (controlled for the maternal genotype) tended to be associated with greater post-natal weight gain than when the C allele was transmitted ($P = 0.006$; by qTDT $P = 0.009$). There appeared to be parent-of-origin effects on post-natal weight gain, with only maternal transmission showing the above association ($P = 0.006$; by qTDT $P = 0.022$). SNPs 34, 38 and 69 failed to show any significant associations with foetal and early growth phenotype.

Comparisons between *INS* VNTR classes I and III subjects as found in non-sub-Saharan-African populations

There was no significant difference between *INS* VNTR classes I and III infants or mothers in birth weight, head circumference at birth or length at birth (all $P > 0.05$). 95% confidence intervals for the infant class I versus class III effect were: birth weight (–22, 38 g), head circumference (–0.7, 1.5 mm) and birth length (–1.2, 4.7 mm).

Discussion

This study is the first that we are aware of finding significant associations in an in situ African population combining polymorphisms in a more genetically diverse sub-Saharan African gene, in this case *INS* and its VNTR, with markers of foetal and early growth. Whilst associations have been variably observed with SNPs that flank the *INS* VNTR in non-African populations, we reasoned that studying an almost exclusively Mandinkan population living in The Gambia would give us an opportunity to search for enhanced associations with purely African polymorphisms.

Recently some of the SNPs and phenotypes that we investigated were studied in African Americans (Adkins et al. 2008), but this population would have been unlike the one we studied due to both the different environmental exposures (nutritional, medical and otherwise) and likely genetic admixture (Parra et al. 1998).

We studied 2 SNPs and 1 indel that were African-specific. All three polymorphisms may have undergone recurrent mutation (or gene conversion) in the past (Stead and Jeffreys 2002), so a given allele at each of these SNPs might be present on very different ancestral lineages. Associations with these polymorphisms therefore have to be interpreted with caution. We also studied a further 3 SNPs that were not African-specific. Of these six polymorphisms we could find no associations between markers of foetal and early growth and SNPs 24, 34, 38 and 69, despite there being evidence that SNP 69 can activate a downstream cryptic 5' splice site to extend the *INS* pre-mRNA 5' leader (Králóvičová et al. 2006), and that SNP 34 in the offspring and SNP 24 in the mother may be associated with risk of being SGA in African Americans (Adkins et al. 2008).

The infant's maternally transmitted (African-specific) SNP 28 allele was associated with post-natal weight gain, an important factor for neonatal survival (Victora et al. 2001). The association was adjusted for the maternal genotype suggesting that it was not caused by maternal genotype effects on the intrauterine environment. The *INS* gene is not considered one of the classically imprinted genes, unlike nearby *H19* and *IGF2*, common polymorphisms in which have been associated with foetal growth in non-African populations (Petry et al. 2005; Kaku et al. 2007). There is evidence that the imprinting of *INS* is developmentally regulated, however, such that only the paternal allele is expressed in the yolk sac (Moore et al. 2001). Our parent-of-origin effect obviously reflects an association with the maternally transmitted allele, however, so the mechanism is unclear although not unique for this gene (Adkins et al. 2008). This may reflect linkage disequilibrium with SNPs in nearby paternally imprinted genes such as *H19*, polymor-

Table 4 Association (modelled with random effects and adjusted for sex of the child, village where the birth took place, mother's parity and season and year of birth) between birth and early growth characteristicsand INS (a) SNP 24, (b) SNP 27, (c) SNP 28, (d) SNP 34, (e) SNP 38 and (f) SNP 69. Results of the statistical analyses are presented as effect size (β coefficient) and 95% confidence interval

Source	Birth length (mm)	Birth head circumference (mm)	Birth weight (g)	Post-natal weight gain (z score points/year)
(a) SNP 24				
Infant genotype	-0.4 (-3.5, 2.8) (n = 664)	-0.4 (-1.5, 0.7) (n = 1,176)	8 (-21, 38) (n = 1,450)	0.02 (-0.02, 0.06) (n = 1,433)
Infant's allele derived from mother controlled for maternal genotype	-5.2 (-14.4, 3.9) (n = 356)	-0.4 (-3.6, 2.8) (n = 601)	2 (-82, 86) (n = 729)	0.08 (-0.04, 0.20) (n = 710)
Infant's allele derived from father	1.3 (-6.0, 8.5) (n = 356)	-0.1 (-2.5, 2.3) (n = 601)	3 (-65, 70) (n = 729)	0.04 (-0.06, 0.13) (n = 710)
Maternal genotype	-0.2 (-3.9, 3.4) (n = 593)	-0.8 (-2.1, 0.6) (n = 1,027)	22 (-18, 62) (n = 1,254)	-0.05 (-0.11, 0.01) (n = 1,235)
Maternal genotype controlled for infant's maternally derived allele	1.3 (-5.5, 8.1) (n = 356)	-1.2 (-3.5, 1.1) (n = 601)	28 (-40, 96) (n = 729)	-0.06 (-0.15, 0.04) (n = 710)
Parent-of-origin	-5.4 (-15.3, 4.5) (n = 356)	-1.5 (-4.9, 1.9) (n = 601)	20 (-79, 120) (n = 729)	-0.00 (-0.15, 0.14) (n = 710)
(b) SNP 27				
Infant genotype	-2.8 (-7.2, 1.6) (n = 667)	-0.4 (-2.0, 1.1) (n = 1,182)	-27 (-69, 15) (n = 1,458)	-0.04 (-0.10, 0.02) (n = 1,441)
Infant's allele derived from mother controlled for maternal genotype	17.5* (5.2, 29.8) (n = 361)	0.0 (-4.0, 4.1) (n = 620)	-25 (-128, 78) (n = 756)	-0.04 (-0.18, 0.11) (n = 736)
Infant's allele derived from father	-6.6 (-15.9, 2.7) (n = 361)	-1.9 (-5.1, 1.3) (n = 620)	-28 (-118, 61) (n = 756)	-0.14 (-0.27, -0.02) (n = 736)
Maternal genotype	-2.0 (-7.4, 3.4) (n = 596)	0.2 (-1.8, 2.3) (n = 1,034)	-14 (-72, 44) (n = 1,263)	0.04 (-0.05, 0.12) (n = 1,244)
Maternal genotype controlled for infant's maternally derived allele	-11.1 (-19.7, -2.4) (n = 361)	-0.7 (-3.7, 2.3) (n = 620)	-17 (-109, 76) (n = 756)	0.04 (-0.09, 0.17) (n = 736)
Parent-of-origin	16.3 (2.0, 30.6) (n = 361)	1.6 (-3.2, 6.4) (n = 620)	-4 (-134, 126) (n = 756)	0.14 (-0.04, 0.32) (n = 736)
(c) SNP 28				
Infant genotype	0.5 (-3.3, 4.3) (n = 658)	-0.2 (-1.5, 1.22) (n = 1,165)	5 (-31, 41) (n = 1,434)	0.02 (-0.04, 0.07) (n = 1,417)
Infant's allele derived from mother controlled for maternal genotype	-6.7 (-16.3, 2.9) (n = 356)	-2.5 (-6.0, 1.0) (n = 610)	-63 (-154, 28) (n = 733)	0.19* (0.05, 0.32) (n = 728)
Infant's allele derived from father	1.0 (-6.7, 8.6) (n = 356)	0.6 (-2.0, 3.2) (n = 610)	18 (-57, 92) (n = 733)	-0.07 (-0.17, 0.04) (n = 728)
Maternal genotype	1.2 (-3.3, 5.6) (n = 595)	-0.4 (-2.0, 1.2) (n = 1,035)	-17 (-66, 31) (n = 1,262)	0.00 (-0.07, 0.08) (n = 1,246)
Maternal genotype controlled for infant's maternally derived allele	4.3 (-2.8, 11.4) (n = 356)	0.7 (-1.9, 3.2) (n = 610)	6 (-71, 83) (n = 733)	-0.06 (-0.16, 0.05) (n = 728)

Table 4 continued

Source	Birth length (mm)	Birth head circumference (mm)	Birth weight (g)	Post-natal weight gain (z score points/year)
Parent-of-origin	−4.1 (−15.0, 6.8) (n = 356)	−2.6 (−6.3, 1.2) (n = 610)	−78 (−187, 31) (n = 733)	0.21* (0.06, 0.37) (n = 728)
(d) SNP 34				
Infant genotype	−3.4 (−7.5, 0.7) (n = 656)	−0.6 (−2.0, 0.9) (n = 1,155)	−46 (−86, −6) (n = 1,429)	−0.01 (−0.07, 0.05) (n = 1,407)
Infant's allele derived from mother controlled for maternal genotype	10.7 (−1.1, 22.5) (n = 352)	0.8 (−3.1, 4.8) (n = 592)	−29 (−132, 75) (n = 725)	0.01 (−0.13, 0.15) (n = 712)
Infant's allele derived from father	−2.3 (−11.1, 6.6) (n = 352)	−0.4 (−3.5, 2.7) (n = 592)	−18 (−105, 69) (n = 725)	−0.13 (−0.25, −0.01) (n = 712)
Maternal genotype	−2.9 (−8.5, 2.7) (n = 581)	0.4 (−1.7, 2.4) (n = 1,012)	−19 (−78, 40) (n = 1,235)	0.04 (−0.05, 0.13) (n = 1,222)
Maternal genotype controlled for infant's maternally derived allele	−13.3* (−22.6, −4.0) (n = 352)	−1.4 (−4.3, 1.7) (n = 592)	−35 (−132, 63) (n = 725)	−0.01 (−0.148, 0.125) (n = 712)
Parent-of-origin	4.5 (−9.2, 18.1) (n = 352)	0.4 (−4.3, 5.1) (n = 592)	−28 (−155, 99) (n = 725)	0.14 (−0.04, 0.31) (n = 712)
(e) SNP 38				
Infant genotype	1.5 (−3.1, 6.1) (n = 666)	−0.2 (−1.8, 1.5) (n = 1,180)	−8 (−53, 37) (n = 1,456)	0.02 (−0.05, 0.08) (n = 1,438)
Infant's allele derived from mother controlled for maternal genotype	15.3 (2.2, 28.3) (n = 373)	−1.4 (−5.9, 3.1) (n = 633)	−57 (−173, 58) (n = 762)	−0.03 (−0.19, 0.14) (n = 748)
Infant's allele derived from father	7.5 (−4.2, 19.2) (n = 373)	0.6 (−3.2, 4.4) (n = 633)	25 (−78, 128) (n = 762)	−0.01 (−0.16, 0.14) (n = 748)
Maternal genotype	1.3 (−4.8, 7.4) (n = 596)	0.4 (−1.8, 2.6) (n = 1,034)	−1 (−64, 63) (n = 1,263)	0.12 (0.03, 0.21) (n = 1,244)
Maternal genotype controlled for infant's maternally derived allele	−1.1 (−12.2, 9.9) (n = 373)	2.6 (−1.0, 6.3) (n = 633)	39 (−65, 143) (n = 762)	0.06 (−0.09, 0.21) (n = 748)
Parent-of-origin	6.4 (−9.4, 22.2) (n = 373)	0.2 (−4.9, 5.4) (n = 633)	−58 (−202, 85) (n = 762)	0.03 (−0.18, 0.24) (n = 748)
(f) SNP 69				
Infant genotype	0.4 (−2.9, 3.6) (n = 661)	−0.6 (−1.7, 0.6) (n = 1,175)	9 (−22, 41) (n = 1,450)	0.01 (−0.03, 0.06) (n = 1,432)
Infant's allele derived from mother controlled for maternal genotype	−0.2 (−9.4, 9.0) (n = 367)	0.1 (−3.0, 3.3) (n = 620)	−5 (−88, 78) (n = 749)	−0.01 (−0.12, 0.11) (n = 730)
Infant's allele derived from father	2.0 (−5.1, 9.2) (n = 367)	−0.1 (−3.5, 1.5) (n = 620)	−11 (−80, 57) (n = 749)	−0.00 (−0.10, 0.09) (n = 730)
Maternal genotype	2.1 (−1.9, 6.1) (n = 596)	−1.0 (−2.5, 0.5) (n = 1,033)	34 (−11, 78) (n = 1,262)	−0.03 (−0.09, 0.04) (n = 1,244)

Table 4 continued

Source	Birth length (mm)	Birth head circumference (mm)	Birth weight (g)	Post-natal weight gain (z score points/year)
Maternal genotype controlled for infant's maternally derived allele	2.7 (−4.6, 9.9) (n = 367)	−2.0 (−4.5, 0.4) (n = 620)	43 (−30, 166) (n = 749)	0.02 (−0.08, 0.12) (n = 730)
Parent-of-origin	−0.4 (−11.5, 10.6) (n = 367)	−0.4 (−4.2, 3.5) (n = 620)	37 (−71, 145) (n = 749)	0.01 (−0.14, 0.16) (n = 730)

* $P < 0.008$

phic variation in which was associated with early weight gain in a British population (Petry et al. 2005).

SNP 27 has been extensively studied in relation to early growth due to its almost perfect linkage disequilibrium with the two major *INS* VNTR classes found in non-African populations (Barratt et al. 2004). In these populations some studies have shown associations with size at birth or early growth (Dunger et al. 1998; Ong et al. 2004; Lindsay et al. 2003; Heude et al. 2006; Osada et al. 2007) and some have not (Mitchell et al. 2004; Bennett et al. 2004; Vu-Hong et al. 2006; Landmann et al. 2006; Mook-Kanamori et al. 2007). This possibly reflects differences in linkage disequilibrium with a putative functional polymorphism between these different populations, although SNP 27 itself might be functional (Královičová et al. 2006). In the current study we could find no association with these phenotypes and the classical VNTR classes or between SNP 27 and either birth weight or head circumference. Where we did find a significant association was between SNP 27 and birth length. This confirms one of our original findings in Caucasians, although the association was in the opposite direction (Dunger et al. 1998), as has previously been observed for the association with birth weight in Pima Indians (Lindsay et al. 2003). People in the current study with the 'T' allele were born shorter at birth with similar weights suggesting that they may have had more body fat. Birth weight associations may therefore not have been evident partially due to poor maternal nutrition during pregnancy.

This is the first time, to the best of our knowledge, where a gene previously found to be in association with a particular quantitative trait that is related to neonatal mortality in non-African populations has been studied in an attempt to find further associations between the more polymorphic African gene and phenotype from an environment with higher neonatal mortalities. Where we found significant associations confirmation in similar populations will be necessary if a suitable population can be found. To interpret our data we applied a Bonferonni correction to account for multiple testing, although this is controversial for correlated data (Rothman 1990). The statistical power for this study, calculated a priori using an effect size based on the one that we originally observed for the *INS* VNTR in a contemporary

British population (Dunger et al. 1998), may not have been realised in practice given that other studies have failed to replicate such an effect size (e.g. Bennett et al. 2004; Mitchell et al. 2004). Nevertheless, given the tight confidence intervals that were gained, suggesting precise group mean estimates, we can be confident that we did not miss any associations with large or moderate effect sizes.

In conclusion this study suggests that in the tribally and socially pure Gambian population that we studied there are associations between SNPs in the *INS* region (one of which is African-specific) and foetal and early growth characteristics, which contribute to the overall polygenic associations with size at birth (Dunger et al. 2007). Despite D' values suggesting a high degree of linkage disequilibrium between the polymorphisms studied, in our population SNPs 34 and 27 were the only two polymorphisms genotyped where there was significant correlation of any extent between genotypes, so associations with different SNPs can not be explained solely by linkage disequilibrium. We could find no further evidence of this genetic region providing a survival advantage when the system was stressed by maternal undernutrition taking place in the rainy season.

Acknowledgments The authors wish to thank the villagers of Keneba, Manduar and Kantong Kunda for their willing participation in this and other studies over the years. Funding for the genotyping part of this study was kindly provided by the Wellcome Trust. Core funding for the prospective study of early growth in the West Kiang District was provided by the Medical Research Council. We also acknowledge the National Institute for Health Research (NIHR) Cambridge Comprehensive Biomedical Research Centre.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Adkins RM, Krushkal J, Klauser CK, Magann EF, Morrison JC, Somes G (2008) Association between small for gestational age and paternally inherited 5' insulin haplotypes. *Int J Obes (Lond)* 32:372–380
- Allison DB (1997) Transmission-disequilibrium tests for quantitative traits. *Am J Hum Genet* 60:676–690

- Barratt BJ, Payne F, Lowe CE et al (2004) Remapping the insulin gene/IDDM2 locus in type 1 diabetes. *Diabetes* 53:1884–1889
- Bennett AJ, Sovio U, Ruokonen A et al (2004) Variation at the insulin gene VNTR (variable number tandem repeat) polymorphism and early growth: studies in a large Finnish birth cohort. *Diabetes* 53:2126–2131
- Carter KW, McCaskie PA, Palmer LJ (2006) JLIN: a java based linkage disequilibrium plotter. *BMC Bioinformatics* 7:60
- Ceesay SM, Prentice AM, Cole TJ et al (1997) Effects on birth weight and perinatal mortality of maternal dietary supplements in rural Gambia: 5 year randomised controlled trial. *BMJ* 315:786–790
- Dubowitz LM, Dubowitz V, Goldberg C (1970) Clinical assessment of gestational age in the newborn infant. *J Pediatr* 77:1–10
- Dunger DB, Ong KK, Huxtable SJ et al (1998) Association of the INS VNTR with size at birth. *Nat Genet* 19:98–100
- Dunger DB, Petry CJ, Ong KK (2007) Genetics of size at birth. *Diabetes Care* 30(Suppl. 2):S150–S155
- Eaves IA, Bennett ST, Forster P et al (1999) Transmission ratio distortion at the INS-IGF2 VNTR. *Nat Genet* 22:324–325
- George V, Tiwari HK, Zhu X, Elston RC (1999) A test of transmission/disequilibrium for quantitative traits in pedigree data, by multiple regression. *Am J Hum Genet* 65:236–245
- Gu D, O'Dell SD, Chen XH, Miller GJ, Day IN (2002) Evidence of multiple causal sites affecting weight in the IGF2-INS-TH region of human chromosome 11. *Hum Genet* 110:173–181
- Heude B, Petry CJ, Avon Longitudinal Study of Parents Children (ALSPAC) study team et al. (2006) The insulin gene variable number of tandem repeat: associations and interactions with childhood body fat mass and insulin secretion in normal children. *J Clin Endocrinol Metab* 91:2770–2775
- Kaku K, Osada H, Seki K, Sekiya S (2007) Insulin-like growth factor 2 (IGF2) and IGF2 receptor gene variants are associated with fetal growth. *Acta Paediatr* 96:363–367
- Královičová J, Gaunt TR, Rodriguez S, Wood PJ, Day IN, Vorechovsky I (2006) Variants in the human insulin gene that affect pre-mRNA splicing: is –23HphI a functional single nucleotide polymorphism at IDDM2? *Diabetes* 55:260–264
- Landmann E, Geller F, Schilling J, Rudloff S, Foeller-Gaudier E, Gortner L (2006) Absence of the wild-type allele (192 base pairs) of a polymorphism in the promoter region of the IGF-I gene but not a polymorphism in the insulin gene variable number of tandem repeat locus is associated with accelerated weight gain in infancy. *Pediatrics* 118:2374–2379
- Lange C, DeMeo DL, Laird NM (2002) Power and design considerations for a general class of family-based association tests: quantitative traits. *Am J Hum Genet* 71:1330–1341
- Lange C, DeMeo D, Silverman EK, Weiss ST, Laird NM (2004) PBAT: tools for family-based association studies. *Am J Hum Genet* 74:367–369
- Lindsay RS, Hanson RL, Wiedrich C, Knowler WC, Bennett PH, Baier LJ (2003) The insulin gene variable number tandem repeat class I/III polymorphism is in linkage disequilibrium with birth weight but not Type 2 diabetes in the Pima population. *Diabetes* 52:187–193
- Lustig O, Ariel I, Ilan J, Lev-Lehman E, De-Groot N, Hochberg A (1994) Expression of the imprinted gene H19 in the human fetus. *Mol Reprod Dev* 38:239–246
- Mathews F, Yudkin P, Neil A (1999) Influence of maternal nutrition on outcome of pregnancy: prospective cohort study. *BMJ* 319:339–343
- McCormick MC (1985) The contribution of low birth weight to infant mortality and childhood morbidity. *N Engl J Med* 312:82–90
- Mitchell SM, Hattersley AT, Knight B et al (2004) Lack of support for a role of the insulin gene variable number of tandem repeats minisatellite (INS-VNTR) locus in fetal growth or type 2 diabetes-related intermediate traits in United Kingdom populations. *J Clin Endocrinol Metab* 89:310–317
- Mook-Kanamori DO, Geelhoed JJ, Steegers EA et al (2007) Insulin gene variable number of tandem repeats is not associated with weight from fetal life until infancy: the Generation R Study. *Eur J Endocrinol* 157:741–748
- Moore GE, Abu-Amero SN, Bell G et al (2001) Evidence that insulin is imprinted in the human yolk sac. *Diabetes* 50:199–203
- Moore VM, Davies MJ, Willson KJ, Worsley A, Robinson JS (2004) Dietary composition of pregnant women is related to size of the baby at birth. *J Nutr* 134:1820–1826
- Ong KK, Preece MA, Emmett PM, Ahmed ML, Dunger DB, ALSPAC Study Team (2002) Size at birth and early childhood growth in relation to maternal smoking, parity and infant breast-feeding: longitudinal birth cohort study and analysis. *Pediatr Res* 52:863–867
- Ong KK, Petry CJ, Barratt BJ et al (2004) Maternal-fetal interactions and birth order influence insulin variable number of tandem repeats allele class associations with head size at birth and childhood weight gain. *Diabetes* 53:1128–1133
- Osada H, Seki K, Sekiya S (2007) Genetic variations within the insulin gene region are associated with accelerated fetal growth. *Tohoku J Exp Med* 212:27–34
- Paquette J, Giannoukakis N, Polychronakos C, Vafiadis P, Deal C (1998) The INS 5' variable number of tandem repeats is associated with IGF2 expression in humans. *J Biol Chem* 273:14158–14164
- Parra EJ, Marcini A, Akey J et al (1998) Estimating African American admixture proportions by use of population-specific alleles. *Am J Hum Genet* 63:1839–1851
- Petry CJ, Ong KK, Barratt BJ et al (2005) Common polymorphism in H19 associated with birthweight and cord blood IGF-II levels in humans. *BMC Genet* 6:22
- Prentice AM, Cole TJ, Foord FA, Lamb WH, Whitehead RG (1987) Increased birthweight after prenatal dietary supplementation of rural African women. *Am J Clin Nutr* 46:912–925
- Rayco-Solon P, Moore SE, Fulford AJ, Prentice AM (2004) Fifty-year mortality trends in three rural African villages. *Trop Med Int Health* 9:1151–1160
- Rothman KJ (1990) No adjustments are needed for multiple comparisons. *Epidemiology* 1:43–46
- Sirugo G, Schim van der Loeff M, Sam O et al (2004) A national DNA bank in The Gambia, West Africa, and genomic research in developing countries. *Nat Genet* 36:785–786
- Stat Corp LP (2007) Stata version 10. StataCorp, Texas
- Stead JD, Jeffreys AJ (2002) Structural analysis of insulin minisatellite alleles reveals unusually large differences in diversity between Africans and non-Africans. *Am J Hum Genet* 71:1273–1284
- Stead JD, Hurler ME, Jeffreys AJ (2003) Global haplotype diversity in the human insulin gene region. *Genome Res* 13:2101–2111
- Stringer C, McKie R (1996) African exodus: the origins of modern humanity. Henry Holt & Co., New York
- Taillon-Miller P, Bauer-Sardina A, Saccone NL et al (2000) Juxtaposed regions of extensive and minimal linkage disequilibrium in human Xq25 and Xq28. *Nat Genet* 25:324–328
- Victoria CG, Barros FC, Horta BL, Martorell R (2001) Short-term benefits of catch-up growth for small-for-gestational-age infants. *Int J Epidemiol* 30:1325–1330
- Vu-Hong TA, Durand E, Deghmoun S et al (2006) The INS VNTR locus does not associate with smallness for gestational age (SGA) but interacts with SGA to increase insulin resistance in young adults. *J Clin Endocrinol Metab* 91:2437–2440
- Williams RL, Creasy RK, Cunningham GC, Hawes WE, Norris FD, Tashiro M (1982) Fetal growth and perinatal viability in California. *Obstet Gynecol* 59:624–632